

OpenAI's smaller, cheaper ChatGPT variant intensifies competition with Anthropic, Google

Article

The news: OpenAI released **GPT-4o Mini**, a smaller and more affordable version of its GPT-4o large language model (LLM).

The multimodal generative AI (genAI) solution can handle text and images and reportedly outperforms **Google's Gemini Flash** and **Anthropic's Claude Haiku** by 5% to 15% in certain benchmarks, per [The Register](#).

AI offerings are diversifying: Understanding that there's no one-size-fits-all approach to genAI solutions, companies are offering smaller models aimed at more focused applications and at a more affordable price.

- Despite strong benchmarks, **OpenAI no longer holds a commanding lead in the LLM market**. Competitors like Anthropic and Google have closed the gap with their own advanced models.
- Other notable mini models in the increasingly crowded LLM space include **Nvidia** and **Mistral's Nemotron-4-340B-Instruct** and **Meta's Llama-3-70B-Instruct**.
- GPT-4o Mini is cheaper than [Claude 3 Haiku](#) and Gemini 1.5 Flash, making it an attractive option for cost-conscious users preferring ChatGPT's functionality.
- OpenAI's Mini supports 128,000 input tokens (images and text) and 16,000 output tokens. Most models handle about 4,000 tokens—**this makes GPT-4o Mini ideal for translation and transformation tasks with large outputs**.

The trend: The launch of GPT-4o Mini underscores a growing trend toward smaller, more cost-effective LLMs. The trend is timely given the arrival of new hardware that can run on-device AI processes efficiently.

- **These models reduce power and cloud computing requirements**, making advanced AI accessible to a broader market.
- This shift is crucial as **it reduces the risk of monopoly in the AI industry**, encouraging competition and providing users with more options.

Our take: Businesses and consumers will increasingly see a range of genAI solutions on a sliding scale of pricing, which makes adoption more affordable across the board and intensifies competition in terms of cost and performance.

State of Generative AI Adoption in Marketing Initiatives, According to B2B and B2C CMOs and Executives Worldwide, Oct 2023

% of respondents



- Using across multiple/few initiatives
- Testing/experimenting
- Not using, but have a strategy
- Not using, and don't have a strategy
- Taking a "wait and watch" approach
- Used in the past, but don't use it now
- Don't know what it is and have no intention to bring it into marketing soon

Note: n=1,800; numbers may not add up to 100% due to rounding
 Source: Capgemini, "Generative AI and the Evolving Role of Marketing: A CMO's Playbook," Dec 11, 2023

284319

